Application Of Dimensionality Reduction Techniques to HRTFs for Interactive Virtual Environments

Bill Kapralos
Business and Information Technology,
University of Ontario Institute of Technology.
Oshawa, Ontario, Canada. L1H 7K4.
bill.kapralos@uoit.ca

Nathan Mekuz
Computer Science and Engineering,
Centre for Vision Research, York University.
Toronto, Ontario, Canada. M3J 1P3.
mekuz@cse.yorku.ca

ABSTRACT

Fundamental to the generation of 3D audio is the HRTF processing of acoustical signals. Unfortunately, given the high dimensionality of HRTFs, incorporating them into dynamic/interactive virtual environment and gaming applications is computationally very demanding. This greatly limits the performance of such applications that incorporate real-time 3D audio. This paper examines the application of data reduction models to HRTFs. In particular, the locally linear Isomap, Locally Linear Embedding (LLE), and the globally linear Principal Components Analysis (PCA) dimensionality reduction tools are applied to the MIT HRTF dataset. Our motivation is to project the inherently high-dimensional space inherent in HRTF measurements onto a lower dimensionality such that they can be incorporated into interactive virtual environments and gaming applications.

Categories and Subject Descriptors

I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Virtual reality*

General Terms

Human Factors

Keywords

HRTF, 3D sound, dimensionality reduction, Principal Components Analysis (PCA), Local Linear Embedding (LLE), Isomap.

1. INTRODUCTION

Collectively, the filtering effects of a sound by the listener's head, shoulders, upper torso, and most notably the pinna, are modeled by a complex response function known as the *head-related transfer function* (HRTF). The HRTF modifies the spectrum and timing of sound signals reaching each ear in a location-dependent manner [1]. Although the HRTF can vary widely amongst individuals, several "generic" HRTF datasets are available, including the one from MIT that contains 710 HRTF measurements (each consisting of a 512-dimensional real-valued vector), measured from an anthropomorphic "dummy head" [2].

ACE'07, June 13–15, 2007, Salzburg, Austria.
ACM 978-1-59593-640-0/07/0006.

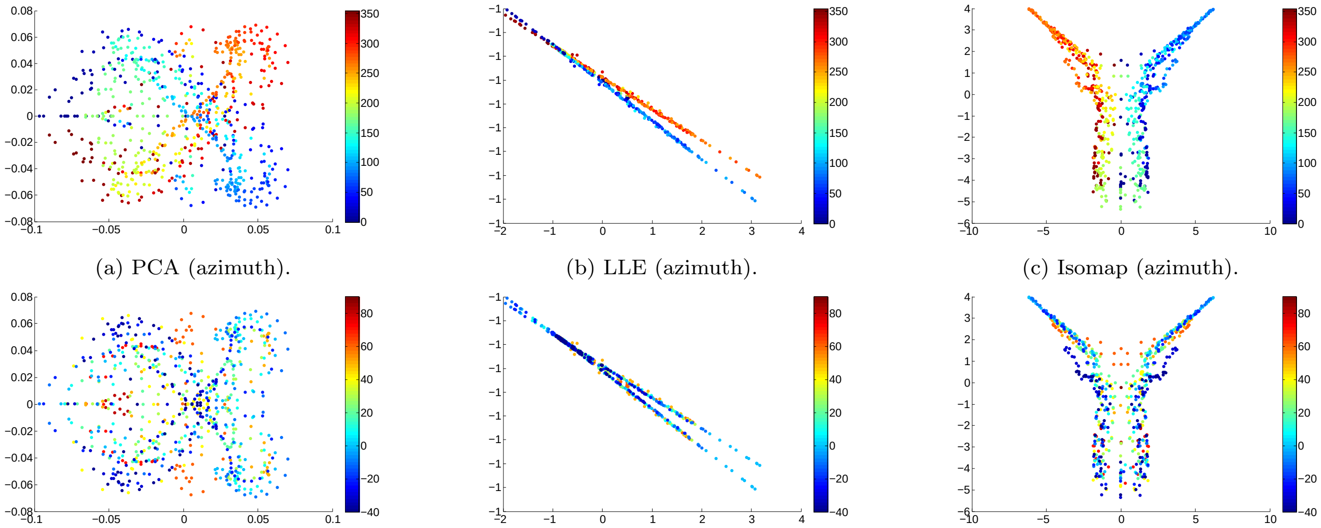
Each measured HRTF forms the basis of a filter that can be used to modulate source sound material (e.g., anechoic sound or synthesized sound). The auditory signal delivered to the left and right ears is obtained by filtering (via a convolution operation) the sound with the coefficients corresponding to the measured left and right ear HRTFs respectively. Presented with the filtered sounds, the user obtains the impression of a sound source at the position corresponding to the measured HRTF (e.g., the desired synthesis location). Convolution is unfortunately an extremely computationally demanding technique thus greatly limiting the performance of any real time 3D audio system.

Dimensionality reduction is a statistical tool commonly used to map high-dimensional data such as images and speech signals into a lower dimensional subspace. The transformed data is typically more suitable for regression analysis or classification than the original data. HRTF data is of very high dimensionality. However, depending on the application, some compression of the HRTF data may be possible. An HRTF compression method that is commonly used is Principal Components Analysis (PCA) [3].

This paper examines the application of several data reduction techniques to the MIT HRTF dataset [2] and evaluates their performance. Our motivation in applying dimensionality reduction to HRTF data is two-fold. In addition to the realized compression, we seek a subspace representation that effectively captures and characterizes the intrinsic structure of the data. In addition to PCA, in this paper we examine the application of the Locally Linear Embedding (LLE) [4] and Isomap [5] dimensionality reduction methods to the MIT HRTF dataset [2]. Regardless the technique used, it is applied once to the complete dataset off-line. This results in a lower-dimensional representation thus allowing the compressed HRTFs to be incorporated into interactive virtual environments and gaming applications where dynamic update rates are required.

2. RESULTS

The suitability of the Isomap and LLE techniques to the reduction of HRTF data was tested by applying the techniques to the MIT HRTF dataset. For comparison purposes, the PCA technique was also applied to the same dataset. Here the resulting subspace of each technique is examined. An indication of how well the subspace represents the underlying structure of the data is obtained by examining how closely the axes of the transformed data correlate with the physical properties that generated the data, in this case azimuth and elevation. A graphical summary of the distribu-



(a) PCA (azimuth). (b) LLE (azimuth). (c) Isomap (azimuth).

(d) PCA (elevation). (e) LLE (elevation). (f) Isomap (elevation).

Figure 1: Distribution of azimuth and elevation in the left channel HRTF data.

Technique	Azimuth	Elevation
PCA (1^{st} dimension)	0.53	0.006
PCA (2^{nd} dimension)	0.113	0.000
LLE (1^{st} dimension)	0.023	0.106
LLE (2^{nd} dimension)	0.046	0.105
Isomap (1^{st} dimension)	0.768	0.000
Isomap (2^{nd} dimension)	0.043	0.006

Table 1: Absolute correlation values.

tion of azimuth and elevation respectively in the left channel HRTF data over the first and second dimensions generated by PCA, LLE, and Isomap is provided in the plots of Figure 1. Although not illustrated due to space restrictions, similar distributions were observed for the right channel HRTF data. The x and y axes of the plots capture the first and second dimensions (components) of the embedding respectively. The 2D PCA embeddings (Figures 1(a),(d)) do not exhibit any clear patterns with respect to azimuth or elevation indicating that although the realized subspace captures some of the variance, it does not capture the intrinsic structure of the HRTF data. The embeddings created by LLE (Figures 1(b),(e)) qualitatively appears to reflect the properties of the HRTF data somewhat better, especially with respect to elevation. The embeddings generated by Isomap (Figures 1(c),(f)) appears to correlate most closely, especially with respect to azimuth where the correlation in the resultant structure can clearly be observed. Note that these embeddings were computed in an unsupervised fashion (i.e. input data was unlabeled). The effective subspace representations computed by LLE and Isomap are solely the result of optimizing their respective projection criteria. Examining the LLE and Isomap embeddings, the LLE embedding seems almost one-dimensional with most observations lying on a diagonal axis. One possible explanation for this is that LLE maps data by combining multiple local optimizations that minimize residuals in each neighborhood. Global structure is only achieved through overlap between neighborhoods making it difficult to tune effectively in this case due to the sparsity of the input space with respect to its dimensionality. The embedding achieved by Isomap appears to be distributing the data in the output subspace more effectively thus taking advantage of both dimensions. A quantitative measure of these results is summarized in Table 1 where the absolute correlation $|r|$ observed between the axes of the embeddings and the physical properties of the HRTF vectors is provided.

3. CONCLUSIONS

In this paper the application of PCA, LLE, and Isomap data reduction techniques to HRTFs has been demonstrated. Given the static nature of HRTF datasets, the overhead associated with applying the data reduction techniques is incurred once during an off-line initialization process. Thus use of the reduced dimensionality data can lead to considerable computational savings. Results indicate that the embeddings created by both LLE and Isomap are superior to PCA, with Isomap exhibiting the largest correlation especially with respect to azimuth. Future work will examine validating the results presented here with user tests to provide insight to any perceptual artifacts such data reduction techniques may introduce.

4. REFERENCES

[1] R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, MA. USA, 1994.

[2] W. G. Gardner and K. D. Martin. HRTF measurements of a KEMAR. *J Acoust Soc Am*, 97:3907–3908, 1995.

[3] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principle components analysis and minimum phase reconstruction. *J Acoust Soc Am*, 91:1637–1647, 1992.

[4] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[5] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.